

AUDITORY AND VISUAL EVALUATION OF FIXED-FREQUENCY EVENTS IN TIME-VARYING SIGNALS

Robert L. Alexander, Sile O'Modhrain, Jason A. Gilbert, Thomas H. Zurbuchen

University of Michigan,
Ann Arbor, MI, 48109
robertalexandermusic@gmail.com

ABSTRACT

This study directly compares the auditory and visual analysis capabilities of participants in a structured data analysis task. This task involved the identification of transient fixed-frequency sinusoid events that were embedded within white noise and noise derived from solar wind time series. It was hypothesized that participants would be able to identify the number of embedded events more quickly and accurately through auditory data analysis than through visual analysis. While visual analysis outperformed auditory analysis overall, additional investigation revealed that auditory analysis outperformed vision in instances where these events were embedded in solar wind data. This task - involving the detection of transient periodic activity occurring within background turbulence - closely mirrors a type of spectral analysis conducted by heliospheric scientists. Additionally, several data examples contained embedded events that were correctly identified through audition while being consistently overlooked through visual inspection. The largest disparity between visual and auditory performance was found in the analysis of white noise spectra that contained no embedded events. In these instances, auditory analysis regularly resulted in the identification of events when none were present; a potential reasoning for these false positives is discussed. The results of this study suggest that the analysis capabilities of each modality may vary based largely on the complexity of the masking stimuli that are present.

1. INTRODUCTION

A growing number of spacecraft instruments are producing observations recorded at higher resolution than ever before [1]. The task of extracting new knowledge from increasingly large and complex data sets necessitates the creation of new tools and analysis methods for effective data mining in the heliospheric sciences. Research has consistently demonstrated that auditory display can play a valuable role in this process, particularly in instances where visualization techniques are inadequate in rendering data sets with high dimensionality [2, 3]. Sonification - the practice of transferring information through (non-speech) audio - offers a method for delivering informational cues through a sensory modality that may otherwise be unoccupied [4].

This study investigates the potential of *audification*, a specific form of sonification through which successive data samples from a continuous time-series are isomorphically mapped onto successive amplitude values of a continuous audio signal. This is the most direct form of sonification, as all data

samples are preserved and spectral features within the original data will emerge as timbral components in the resulting audio-file. At the standard rate of sound-file playback, audification enables a listener to extract spectral content from an audio file at a rate of 44,100 samples per second, i.e. to identify spectral components potentially at frequencies up to approximately 20kHz. This type of auditory data representation is ideal for large time-series data sets, and analytical listening can potentially reveal features that may be overlooked by other sensory modalities [5].

The goal of this study is to directly compare the auditory and visual analysis capabilities of participants in a structured data analysis task. The participants consisted of two groups at the University of Michigan who have experience working with spectral displays: heliospheric researchers and computer-music specialists. Transient sinusoidal waveforms were embedded in time-varying signals that contained background noise, and the task of the participant was to identify how many of these transient events occurred within each example. This is similar to a type of spectral analysis task found in the heliospheric sciences. It is hypothesized that participants will be able to identify the number of time-varying fixed-frequency sinusoid events more quickly and effectively through auditory display than through visual analysis. This paper will provide a psychoacoustic context for this study before presenting the experimental design and significant findings. The results will be discussed, and finally, various avenues for future investigation will be proposed.

1.1. Previous research

The current study is a part of a larger investigation into the application of audification for exploratory data analysis within the space sciences, where data visualization techniques have been the standard approach. While the visual modality is extremely powerful, it does have well-known constraints. The relationship between peripheral and focal awareness in vision is somewhat different to that in audition. Whereas visual attention directs the eyes to focus on a particular part of the visual field, auditory attention is guided toward individual 'streams' of sound which are identifiable by the spatially and/or temporally coherent behavior of sound-producing events in the environment [6]. This constraint can be addressed by engaging multiple modalities with different types of information [7]. In this way, the bandwidth of information that is accessible to an analyst is increased, along with the rate of information transfer [8-10]. Early research demonstrated that known visual-analysis methods may often be inferior to auditory display in the

representation of multivariate data [11]. As a diagnostic tool, audification has already proven successful in the evaluation of data generated by the Solar Wind Ion Composition Spectrometer, and it has produced new insight as to the source regions of the solar wind [2, 12].

Successful applications of audification in exploratory data analysis must be coupled with a systematic investigation of the working environment in which this analysis is conducted, along with the psychoacoustic principles that mediate auditory perception. This is necessary in order to establish a set of generalizable knowledge and best practices that can be applied across diverse scientific disciplines. Toward this end, a preliminary goal of this research was to determine the baseline ability of untrained listeners to discriminate between different types of audified data sets and digitally manufactured noise. It was found that pre-exposure to audified data had a strong positive correlation with participants' performance on this task. Additionally, participants with no previous exposure to audified data were able to discriminate between audified scientific data sets and digitally generated noise and sinusoidal waveforms at a rate much higher than chance performance [13].

This study builds on the research of Pauletto and Hunt (2005). Their work uncovered a strong correlation between observations made through auditory and visual analysis methods in the evaluation of complex time-series data [14]. Here it was demonstrated that data analysts are able to make similar assessments through the use of auditory and visual analysis processes, an important first step in establishing equal footing for audification in the evaluation of scientific data sets (as data analysis techniques have heretofore been dominated by vision). This study extends their work in order to determine specific perceptual strengths and weaknesses of auditory display in comparison to visual analysis methods.

1.2. A psychoacoustic investigation

The branch of science that concerns itself with human auditory perception is known as psychoacoustics, and an understanding of the human auditory system is critical in the optimization of auditory displays. The auditory system is able to detect minute changes in pressure, and these changes are broken down into a series of component frequencies in a process akin to a Fast Fourier Transform (FFT), as frequencies ranging from low (20Hz) to high (20kHz) are displaced tonotopically across the basilar membrane. This understanding of audition, known as "place theory," is commonly credited to Hermann Helmholtz [15].

The process of evaluating audified data sets for subtle spectral features can be assessed within Bregman's perceptual framework of auditory scene analysis. The task of the listener is to identify and segregate auditory objects (which Bregman terms 'auditory streams') within a complex auditory scene [16]. Data derived from the solar wind is highly variable and turbulent in nature, and the audification process results in a time-varying signal that is largely 'noisy.' We can utilize the principles of grouping from Gestalt psychology to gain a sense for how the auditory system creates higher-level organization from this complex stimulus (these principles state that we tend to group objects based on proximity, similarity, closure, good continuation, common fate, and good form) [17]. The listener parses the incoming sensory information to determine which

portions of a complex spectrum belong to a single unified auditory object. This study directly compares the abilities of the listener to parse complex spectral content both auditorily and visually [18].

It is important to consider the possibility of subjective masking when establishing parameters for an auditory display. Certain situations may arise in which one piece of audified data could subjectively overshadow another, such that the listener would be unable to perceive both data simultaneously. In this situation, one sound has effectively masked the presence of the other. Masking can occur when two stimuli are presented in a relatively close frequency range, and one is measurably louder than the other. Temporal masking can occur when one sound occurs with an onset time extremely close to a second sound [15]. In the presentation of spectra derived from solar wind time-series, it is possible that this masking could occur both visually and/or auditorily, as a broadband visual stimulus may effectively mask the presence of a second stimulus with lower amplitude.

This research is an early step in a systematic exploration of the underlying perceptual phenomena that mediate the multi-modal data analysis process. Exploratory data analysis involves the evaluation of complex data sets for the presence of underlying patterns and structures [16]. Through data audification, the auditory system will begin to segregate meaningful information from complex background noise through the application of selective attention; a function commonly referred to as the "Cocktail Party Effect" [19]. It has been suggested that audification can be extremely effective in the detection of equipment-induced noise [3], which can manifest itself in any number of ways, from broadband spectral distortion to discrete periodic components. This is one example in which distinguishing meaningful spectral properties from background noise becomes a particularly important task.

1.3. Origins of the analysis task

In the case of this study, the meaningful stimuli (fixed-frequency sinusoids) were embedded in a masking signal derived from either white noise or solar wind turbulence, such that auditory and visual performance might be assessed in the presence of varying levels of distractor stimuli. Additionally, the latter case closely resembles an analysis task that a heliospheric research scientist might encounter in the field, as these transient bursts of sinusoidal activity closely mirror several wave modes (e.g., whistler modes and ion cyclotron waves) that can be found in high-resolution magnetometer observations of solar wind turbulence. These waves are of interest to the scientific community because they effectively interact with particles; however, they are often very transient in nature and difficult to identify through traditional analysis methods due to both the turbulent nature of the solar wind and the large volumes of available data.

An extremely clear example of one such event occurred in WIND magnetometer data during June 2008, which is displayed in **Figure 1**. Here, a spectrogram representation is presented that spans roughly 83,000 audified data samples derived from WIND magnetometer observations. Broadband turbulence is manifested as vertical lines, while wave activity is apparent as a single bright object at the center of the spectral display. This is

one particularly clear example; most instances are extremely subtle in comparison.

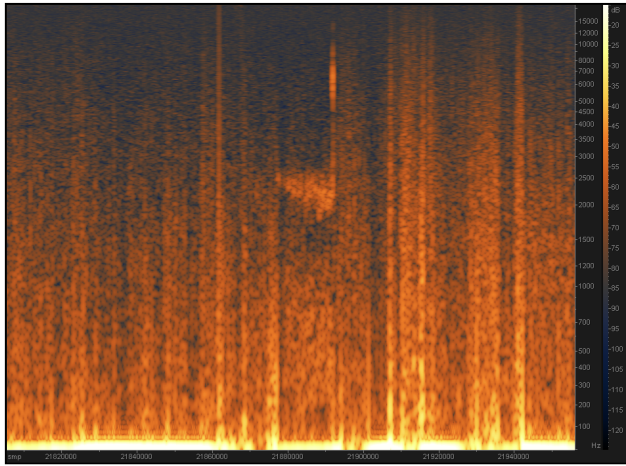


Figure 1: The spectrogram display (reduced in size) of a coherent wave event occurring in high-resolution WIND Magnetometer data during June 2008. This event spans roughly 23 minutes in the original data and 350ms in the resulting audio file.

2. EXPERIMENTAL METHOD

One important guiding question has been: what baseline metrics can be established for auditory display through audification, and how do they compare to visual analysis capabilities? Towards this end, this study directly compares the analysis capabilities of participants who both listened to and viewed data as part of a structured feature identification task. Transient sinusoidal waveforms were embedded in time-varying signals that contained broadband noise. The task of the participant was to identify how many of these transient events occurred within each example. The embedded sinusoidal events were tightly parameterized such that deeper investigation might provide some insight as to the performance of the two modalities in the identification of stimuli with varying amplitude, frequency, and duration. This kind of baseline evaluation is critical in order to gain a deeper understanding of how auditory perception may be applied to complex data analysis tasks, and ultimately integrated into the exploration of large data sets within the sciences.

2.1 Hypothesis

Participants will be able to identify the number of time-varying fixed-frequency sinusoid events more quickly and accurately through auditory data analysis than through visual analysis. Here, accuracy is a comparative measure of the number of events reported by the participant for each example versus the number of events that were actually embedded; this measure provides a margin of error.

2.2 Participants

Ten participants took part in this research study. Half were members of the Solar and Heliospheric Research Group (SHRG) at the University of Michigan, and the other half were

computer-music specialists; all had experience working with spectrogram displays.

2.3 Stimuli

All solar wind data utilized in the study were gathered from magnetometer observations on the ACE and WIND satellites. These time-series data sets were converted to audio files using an audification code written in Matlab. All data samples from the original data sets were preserved in this isomorphic mapping process. All visual stimuli were then rendered in the iZotope RX software environment, and consistent settings were utilized to ensure uniformity across examples. Spectrograms were presented with a linear scaling on the y-axis, and a chromatic color mapping from black representing the absence of energy to white representing the full presence of spectral energy. This visualization method was reviewed by members of the SHRG and deemed appropriate for the spectral representation of solar wind data. **Figure 2** is an example derived from measurements of the magnitude of the solar magnetic field as observed by the WIND satellite over a period spanning August to September 2004.

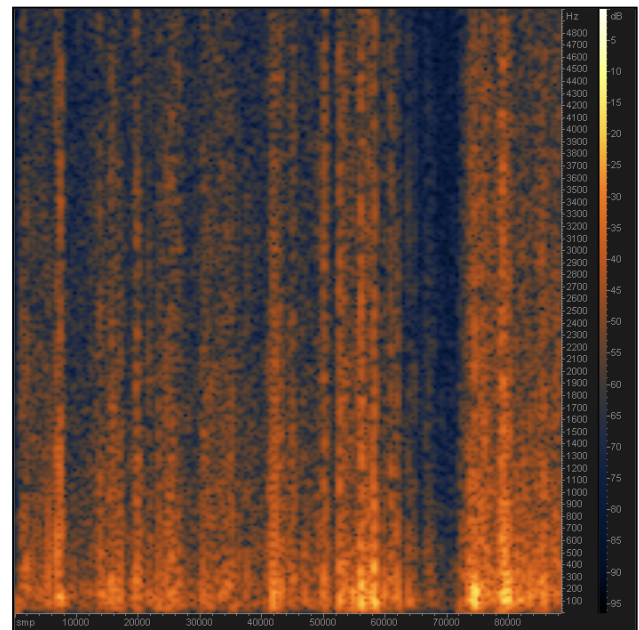


Figure 2: Spectral representation of audified solar wind turbulence. Broadband turbulence is represented as vertical bands of increased brightness.

The fixed-frequency sinusoid events were created with a synthetic data generation module constructed in the Max/MSP computer-music programming environment. These events ranged in frequency from 300Hz to 4.7kHz; intensity varied between -16db, -19db, and -22db (all masking noise was balanced to an RMS level of 0db); and length varied between 25ms 50ms, 100ms, and 200ms. The loudness level, frequency, and duration were held constant within each example, and varied between examples. The number of fixed-frequency events embedded in each example ranged from 0 to 3. All

possible permutations of the data parameters were utilized to create a set of 48 unique stimuli that were embedded in both white noise and solar wind data, resulting in a total of 96 examples. These noise elements acted as masking signals with varying level of complexity. The solar wind data sets were pre-screened in order to minimize the likelihood that they inherently contained any significant fixed-frequency events that may be identified in the auditory and visual analysis tasks.

All examples contained approximately 88,000 data points, which translated to two seconds of audio playback at a sampling rate of 44.1kHz. Four duplicate examples were included for the purpose of confirming internal consistency. All 100 examples were presented to both the auditory and visual modalities in an order that was randomized before the tasks began. The ordering of the two analysis tasks was also randomized across all participants, such that some completed the visual analysis module before moving onto the auditory analysis section, and vice-versa.

2.4 Apparatus

The experiment was conducted on a 15-inch MacBook Pro with the Mac OS X (10.8.2) operating system. The listening task was completed with Audio-Technica ATH-M50 stereo headphones. The pre-test, analysis tasks, and post-test were all encapsulated within a single standalone application constructed with the Max/MSP computer-music programming environment (version 6.05). All responses were recorded using the “coll” object and saved as data files in .txt format. A time-stamp for individual responses was recorded, along with total completion time for each task. Before beginning the experiment, participants were prompted to provide their first name, middle initial and last name; unique 3-letter file names were created from the initials.

2.5 Procedure

Participants were trained to visually and auditorily assess for the presence of fixed-frequency sinusoid events that were embedded in both white noise and noise generated from solar wind data sets. All visual stimuli were presented as spectrogram displays, and auditory stimuli were presented through audification and played back over headphones. These examples were presented sequentially, and participants were not allowed to go back and change their responses once an answer had been provided. The participants’ task was to assess each example, and to report the number of fixed-frequency events they were able to detect. During the analysis task, participant responses were entered into a number box that allowed any integer values between 0 and 99. These values could be entered either by clicking and dragging on the number box, or typing on the keyboard.

One training module guided participants through the process of listening to auditory data, and the other provided assistance in conducting a visual analysis of a spectrogram. These modules both explained the analysis task and guided participants through the interface (the visual training module is displayed in **Figure 3**). The data files used for the training sessions first demonstrated the fixed frequency sinusoids in isolation before introducing the full range of examples that participants would be expected to identify. These data examples were generated specifically for the training task, and were not included in the study. Additionally, participants were not able to

change the volume setting once they completed the auditory training module; this prevented a perceptual bias that could be introduced by a global volume change partway through the experiment.

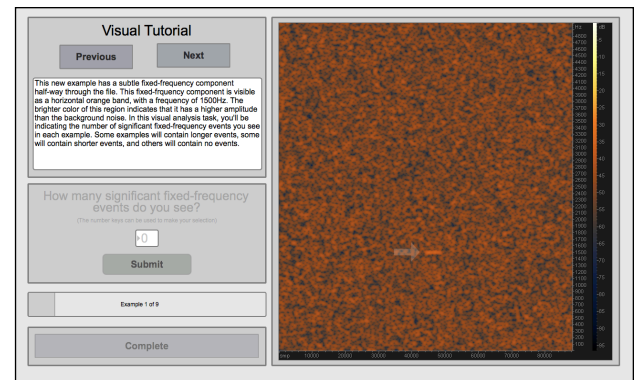


Figure 3: The visual training module provided participants with clear examples of what types of stimuli could be presented in the study.

An exact definition of “fixed-frequency” events was provided in training modules for both modalities, along with examples that demonstrated the types of events participants would be expected to find. All participants reported that they considered the training provided for the analysis task to be easy to understand and/or adequate.

For the auditory portion of the task, participants were provided with several options for starting and stopping playback. An on-screen play-bar could be used to start the sample from any specific location, and the space bar could be used to start and stop playback. Additionally, a looping option allowed participants to listen to the audio repeatedly. For both the auditory and visual portion of the test, a small temporal gap was inserted between the presentation of each example in order to minimize the impact of subtle differences that may be present between subsequent stimuli.

Testing was conducted at various locations at the University of Michigan. All participants completed the analysis tasks in a quiet space that was free from potential distractions. The experiment was administered with an interface constructed in the Max/MSP programming environment. Subjects were provided with headphones and given a brief verbal overview of the task. After completing a short pre-test questionnaire, participants were randomly assigned the visual or auditory task. Participants were informed that while there was no time limit for this test, the total completion time was recorded, and they should attempt to respond “both quickly and accurately.” In order to minimize the effects of fatigue, participants were informed that they could take a short break between the visual and auditory analysis tasks.

The post-test questionnaire was specifically designed to determine the participant’s familiarity with sonification, experience working with spectrograms, level of comfort with computers, and experience with data analysis, mathematical modeling, and scientific research. Participants were asked to rate the difficulty of the listening task in relation to the visual task. This information was gathered in order to assess for a potential correlation between individual backgrounds and

performance on the analysis task. Participants were also asked if they noticed any duplicate audio and/or visual stimuli, and if yes, to write how many. While the duplicate stimuli could be used as a measure of internal consistency, the participant's awareness (or lack thereof) of these duplicate stimuli could also yield potentially valuable insight as to their cognitive state during the examination. Finally, a space was provided for additional feedback in free-response form.

3. RESULTS

In all instances, statistical significance was calculated through the implementation of a matched, 2 tailed t-test. For this study, significance was considered at a value of $p < .05$, and strong significance at a value of $p < .01$. Overall, participants provided correct responses for 66% of the visual stimuli, and 60% of the auditory stimuli, this difference of 6 percentage points was found to be statistically significant ($p < 0.01$). For examples in which fixed-frequency events were embedded in white noise, participants provided correct responses for 66% of the visual stimuli and 54% of the auditory stimuli ($p < 0.01$). For examples in which fixed-frequency events were embedded in noise generated from solar wind data sets, participants provided correct responses for 65% of the visual stimuli and 66% of the auditory stimuli ($p = .94$). A summary of task performance has been provided in **Figure 4**, and additional information has been provided as to performance with the white noise and solar wind data maskers.

		Vis	Aud	<i>p</i> value
Percentage of Correctly Identified Examples	White Noise	66	54	< 0.01
	SW Data	65	66	$= 0.94$
	Average	66	60	< 0.01
Average Error (Per Example)	White Noise	0.55	0.76	< 0.01
	SW Data	0.57	0.48	$= 0.01$
	Average	0.56	0.62	$= 0.03$
Total Number of Events Reported	White Noise	53	68	< 0.01
	SW Data	47	67	< 0.01
	Total	100	135	< 0.01
Average Completion Time (mm:ss)	Average	11:45	18:00	< 0.01

Figure 4 A comparison of auditory and visual performance on the structured identification task. Here "SW" stands for Solar Wind.

The margin of error was calculated by first determining the difference between the participant responses and the correct response for each example, and then averaging across the total number of examples. Participants had an average error margin of 0.56 in the visual analysis task, and 0.62 in the auditory analysis task ($p = 0.03$). For examples in which fixed-frequency events were embedded in white noise, participants had an error margin of 0.55 in the visual task and 0.76 in the auditory task ($p < 0.01$). For examples in which fixed-frequency events were

embedded in noise generated from solar wind data sets, participants had an error margin of 0.57 for the visual stimuli and 0.48 for the auditory stimuli ($p = 0.01$).

The average completion time on the visual analysis task was 11 minutes and 45 seconds; this was 53% faster than the average completion time for the auditory task, which was approximately 18 minutes. On average, participants reported detection of 100 total events in the visual analysis task and 135 events in the auditory analysis task (a 35% increase).

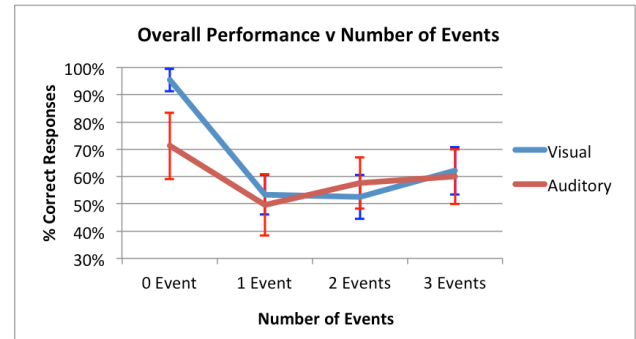


Figure 5: Task performance as a function of the number of embedded fixed-frequency events.

A summary for overall task performance as a function of the number of embedded events is provided in **Figure 5**. Participants utilizing visual analysis correctly identified examples without any embedded stimuli at an average success rate of 96%, this rate was 71% in the auditory analysis task. Respectively, average rates for the successful identification of single events were 53% and 50%; double event identification rates were 58% and 60%; and successful triple event identification rates were 62% and 60%. The difference in performance between the two modalities was only statistically significant for examples containing no events ($p < 0.01$).

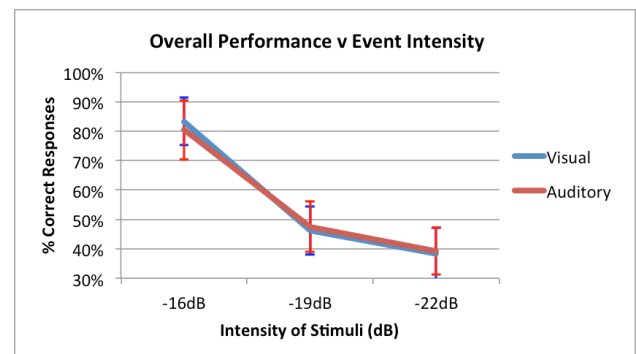


Figure 6. Task performance as a function of stimuli intensity.

A summary for the overall task performance as a function of the intensity of embedded events is provided in **Figure 6**. Performance generally declined as the intensity of embedded events decreased. Events provided at -16dB were visually identified with a success rate of 83%, and auditorily identified at a success rate of 80%. These respective values for events that were presented at -19dB were 46% and 48%; for events presented at -22db, successful identification rates dropped to 38% and 39%. These differences were not found to be statistically significant.

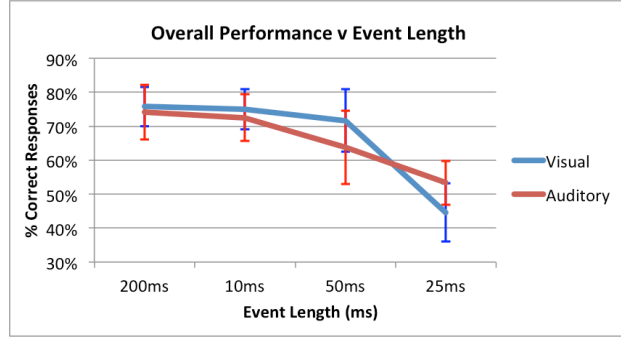


Figure 7: Performance on the identification task as a function of decreasing event length.

A summary of the overall task performance as a function of the length of embedded events is provided in **Figure 7**. Stimuli that contained embedded events with a duration of 200ms were correctly identified visually at an average success rate of 76%, and correctly identified auditorily at an average success rate of 74%. The respective success rates for stimuli containing events with a duration of 100ms were 75% and 73%; success rates dropped to 72% and 64% for events lasting 50ms in duration ($p = 0.018$), and 45% and 53% for events with a duration of 25ms ($p = .049$).

4. DISCUSSION

When performance across all 100 examples is assessed, vision outperformed audition by a margin that was statistically significant ($p < .01$). However, the visual modality was consistently outperformed by the auditory modality in the detection of fixed-frequency sinusoid events embedded in solar wind data sets ($p = .014$). In this case, the average visual identification success rate was 52%, while the auditory success rate was 63%. **Figure 8** provides a summary of the performances of the two modalities in identifying events embedded in solar wind data sets at various levels of intensity. Here it can be seen that the success rate for auditory recognition improved slightly in relation to visual recognition as event intensity declined. Participants who utilized the auditory modality correctly identified 50% of examples that contained events embedded at -22dB, while participants who utilized visual analysis successfully identified 40%. This difference of 10 percentage points was found to be statistically significant ($p = 0.01$).

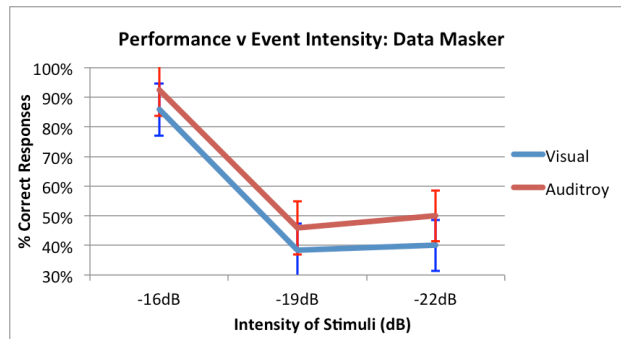


Figure 8: Percentage of correctly identified examples containing fixed-frequency events embedded in solar wind data as a function of the intensity of embedded events.

4.1 Analysis of demographic influence

Demographic information and previous experience (as assessed by the pre- and post-tests) significantly contributed to task performance in many instances. Correlation between demographic information and task performance was determined by calculating Pearson's product-moment correlation coefficient [20]. Equation (1) is used to calculate linear dependencies between two variables, where r is the sample correlation coefficient and x and y are the variables under test. A resulting correlation factor of $r = 1$ indicates a perfect correlation between the two variables, a factor close to zero indicates very little or no correlation, while a factor of -1 indicates a perfect anti-correlation.

Correlation factor

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \quad (1)$$

A strong positive correlation was found between task performance and educational level, with performance increasing as a function of number of years in higher education ($r = 0.83$). A moderate correlation was found between successful identification rates on the two tasks ($r = 0.63$). A moderate negative correlation was found between performance on the two tasks and amount of musical training ($r = -0.46$), as well as a moderate negative correlation between task performance and experience with sound editing and audio processing ($r = -0.66$). One participant found the visual analysis task to be more difficult than the auditory task; five considered the identification task to be easier. Perceived difficulty had no statistically significant effect on task performance.

4.2 False positives in the auditory identification task

While the visual modality had a higher overall success rate in identifying the number of embedded fixed-frequency components, the data reveal several pieces of insightful information upon closer inspection. When the stimuli containing zero events are removed from the analysis, the overall success rate in both modalities evens out at 56%. It is immediately clear that the auditory modality was predisposed towards indicating false-positives in the absence of embedded frequency components. The distribution of these false positives was not completely uniform, multiple examples were labeled as containing one or more embedded events by half of the participants, while other examples were either correctly identified by all participants or incorrectly identified as a false positive by a single participant. White noise is not a truly randomized distribution, and it is possible that upon repeated listening some participants began to pick up on structures occurring at very small time scales.

One participant noted that they attempted to carefully fine-tune their auditory threshold for event detection, and that they "didn't include some short frequency bursts that may be audible." This suggests that they were indeed able to hear short transient fixed-frequency events occurring more rapidly than the 25ms threshold. Another participant noted that they "heard more happening (in the audio file)... This was good but also leads me

to wonder if I had some false positives.” The eye had comparatively little trouble detecting events presented in the relatively uniform visual background created by white noise, however, the non-uniform broadband energy bursts present in the solar wind data sets may have provided a significant amount of visual distraction.

4.3 Closer investigation of individual stimuli

There were several examples utilizing solar wind data that were consistently assessed incorrectly through visual analysis, and correctly assessed through audition. One such example was missed by every participant in the visual task and only missed by one participant auditorily. This example contained a 25ms event at -22db that occurred very close to the end of the data set. Another example that contained a slightly longer event very close to the end of the file was correctly identified only once visually, and eight times auditorily. This indicates that there may potentially be some visual bias away from events that take place at the very edge of a spectrogram.

One example was incorrectly identified seven times visually and only once auditorily. The fixed-frequency element occurring roughly half way through the file at 1.5kHz was subtle, but easily recognized auditorily with training. A second fixed-frequency component, beginning half way through the file, was almost completely visually obscured by the broadband noise event that occurred at the same time. While the same event was also masked by the broadband noise element in the auditory representation, it seems that the ear may not have had as much trouble separating the sinusoidal signal from the background noise.

4.4 Review of experimental design

This study utilized a relatively small pool of ten participants. The recruitment of a larger participant pool was hindered due to the lack of available individuals with the specialized knowledge necessary to complete the identification task. Ideally, future research should work with a larger sample size in order to better determine statistical significance. The use of participants with domain-specific knowledge limits the transferability of these results, as the performance may vary in the general public.

4.5 Evaluating internal consistency

For each participant, four examples were presented twice in each modality in order to determine whether participants' responses were consistent across multiple exposures (the majority of participants indicated that they detected the presence of repeated stimuli when asked in the post-test). No participant was entirely consistent across the four repeated stimuli, and on average participants were consistent in their evaluation of approximately 3 of the 4 stimuli for both modalities. Three participants answered consistently across all repeated visual examples, and three separate participants achieved perfect consistency auditorily. This lack of complete internal consistency indicates that participant evaluations varied slightly over time, which could be attributed to factors such as learning or fatigue, which might improve or degrade performance over time respectively.

This speaks to the difficulty of the analysis task, which required participants to assess for the presence of extremely subtle features. In light of this fact, the lack of complete internal consistency is to be expected, and the effects of learning and fatigue were minimized through both the randomization of the task ordering across participants, and the randomization of the stimuli presented within these tasks.

4.6 General Discussion

Though both identification tasks engaged separate modalities, they were fundamentally similar in that each involved the identification of pre-defined objects embedded within background noise. Visually, these objects were defined by dimensions of color, brightness, length, width, and height; while auditorily they were defined by the frequency space they inhabited, their relative amplitude, and duration. Placed within the context of gestalt theory [21], it could be said that participants utilized these unique properties in establishing, for example, “belongingness” for an explicit subset of the incoming sensory stream. The results of this study, generally speaking, provide some information about the relative ability of the visual and auditory modalities to segregate meaningful information from background noise in the evaluation of certain scientific data sets.

Visually, the spectrogram display of white noise presented a relatively uniform background characterized by a lack of remarkable structures, and this visually unified pattern could be perceptually encoded as a single object against which the embedded features were readily identifiable. Conversely, the spectrogram display of the solar wind spectra contained features on both micro- and macro-scales. The difference between visual and auditory performance when features were embedded in solar wind spectra suggests that the visual modality was comparatively more affected by the presence of complex distractor stimuli than audition, and that this effect was greater in the identification of subtle features. Additionally, the discrepancy between visual and auditory performance in the identification of stimuli embedded in a synthetic noise mask suggests that future research should test the original hypothesis with noise other than white (e.g., pink).

It could be said that the mechanisms that promote auditory stream segregation were brought to bear as participants listened to sounds derived from solar wind data sets, and the auditory system was comparatively more successful in using subtle spectral cues to parse meaningful signals from background noise [16]. This points toward the types of features that may be best suited for recognition through audification – namely those which subtly present themselves within a complex time-varying signal. While visual performance surpassed audition in the identification of sinusoidal events that were 50ms in length, auditory analysis yielded a higher success rate in the identification of the shortest events.

It is common practice for many heliospheric scientists to create visual representations that average a power spectrum over a large number of data samples, and in these instances audification could provide new information regarding the small-scale features that are lost in this process. In this type of practical data analysis task the strengths of one modality may support the weaknesses of another, as audification may reveal subtle spectral features overlooked through visual assessment,

while vision may assist in ruling out events that are too subtle to warrant additional investigation.

A future study should investigate the ability of the auditory and visual modalities when applied in tandem towards a specific scientific data analysis task. In this way, some insight could be gleaned as to how audification compares with visual analysis techniques in real-world scenarios. Additionally, this would shed light on the types of features that are readily identifiable through auditory analysis. An interface such as iZotope RX is an ideal platform for conducting such work, as it provides real time feedback both visually and auditorily, and annotations may be added directly to the data in the form of markers. While this study employed highly parameterized artificially generated stimuli in order to extract some quantitative information as to the relative performance of the two modalities, future research should draw example stimuli from raw data sets as found in the field, and participants could be provided with a more open-ended identification task.

Finally, it is worth noting that, for many of the participants, this was the first instance in which they had utilized auditory analysis in a data analysis task, while some had worked with spectrogram displays for well over a decade. For this reason, it would also be valuable to study the effects of training on participant performance in a set of structured analysis tasks.

5. CONCLUSION

This study directly compared the performance of participants utilizing auditory and visual analysis methods in a structured data analysis task. While visual analysis outperformed auditory analysis overall, additional investigation revealed that auditory analysis outperformed vision when events were embedded in solar wind data as opposed to white noise. In these instances, the identification task closely resembled a type of analysis conducted by heliospheric scientists. When provided with examples that contained no embedded fixed-frequency events, participants utilizing the auditory modality were more likely to report false positives, and it was suggested that this could be attributed to the extreme sensitivity of the auditory modality. Finally, several data examples contained embedded events that were correctly identified through audition while consistently overlooked through visual inspection. These findings support earlier research that revealed a high correlation between assessments made through auditory and visual analysis methods, and further suggest that the analysis capabilities of each modality may vary based largely on the complexity of the masking stimuli that are present.

References

- [1] Board, S.S., *Solar and Space Physics: A Science for a Technological Society*. 2013: National Academies Press.
- [2] Alexander, R.L., Gilbert, J. A., Landi, E., Simoni, M., Zurbuchen, T.H., Roberts, D. A. *Audification as a Diagnostic Tool for Exploratory Heliospheric Data Analysis*. in *Proc. 17th Int. Conf. Aud. Disp (ICAD 2011)*. 2011. Budapest, Hungary.
- [3] Hayward, C., *Listening to the Earth Sing*, in *Auditory display: Sonification, Audification, and Auditory Interfaces*, G. Kramer, Editor. 1994, Addison-Wesley: Reading, Mass. p. 369-404.
- [4] Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N.; Neuhoff, J., Bargar, R., Barrass, S., Berger, J., Evreinov, G., Fitch, W., Gröhn, M., Handel, S., Kaper, H., Levkowitz, H., Lodha, S., Shinn-Cunningham, B., Simoni, M., Tipei, S., *Sonification report: status of the field and research agenda*. 1999: International Community for Auditory Display.
- [5] Florian Dombois, G.E., *Audification*, in *The Sonification Handbook*, A.H. Thomas Hermann, John G. Neuhoff, Editor. 2011, COST Berlin GmbH. p. 564.
- [6] Brock, D. and B. McClimens. *To What Extent do Listeners Use Aural Information When it is Present?* in *International Conference on Auditory Display*. 2011. Budapest.
- [7] Keith, V.N. and S. Barass, *Finding Trading Patterns in Stock Market Data*, B. Stephen, Editor 2004. p. 45-55.
- [8] Pollack, I. and L. Ficks, *Information of Elementary Multidimensional Auditory Displays*. Vol. 26. 1954: ASA. 155-158.
- [9] Fidell, S., *Sensory Function in Multimodal Signal Detection*. Vol. 47. 1970: ASA. 1009-1015.
- [10] Loveless, N.E., J. Brebner, and P. Hamilton, *Bisensory presentation of information*. *Psychological Bulletin*, 1970. **73**(3): p. 161-199.
- [11] Yeung, E.S., *Pattern recognition by audio representation of multivariate analytical data*. *Analytical Chemistry*, 1980. **52**(7): p. 1120-1123.
- [12] Landi, E., Alexander, R.L., Gruesbeck, J.R., Gilbert, J.A., Lepri, S.T., Manchester, W.B., and Zurbuchen, T.H., *Carbon Ionization Stages as a Diagnostic of the Solar Wind*. *The Astrophysical Journal*, 2012. **744**(2): p. 100.
- [13] Alexander, R.L., O'Modhrain, S., Gilbert, J.A. Zurbuchen, T.H., Simoni, M., *Recognition of Audified Data in Untrained Listeners*, in *International Conference on Auditory Display*, B.N.W. Michael A. Nees, Jason Freeman, Editor 2012: Georgia Institute of Technology. p. 135-140.
- [14] Pauletto, S. and A. Hunt. *A Comparison of Audio and visual Analysis of Complex Time-Series Data Sets*. in *International Conference on Auditory Display*. 2005. Limerick, Ireland.
- [15] Moore, B.C.J., *An introduction to the psychology of hearing (5rd ed.)*. 2004, Great Britain: Elsevier Academic Press. xvi, 350.
- [16] Bregman, A., *Auditory Scene Analysis: The Perceptual Organization of Sound*. 1994: A Bradford Book.
- [17] Bregman, A.S., *The formation of auditory streams*. *Attention and performance*, 1978. **7**: p. 63-76.
- [18] Rock, I., *Indirect perception*. 1997: Mit Press.
- [19] Cherry, E.C., *Some Experiments on the Recognition of Speech, with One and with Two Ears*. *The Journal of the Acoustical Society of America*, 1953. **25**(5): p. 975-979.
- [20] Lee Rodgers, J. and W.A. Nicewander, *Thirteen ways to look at the correlation coefficient*. *The American Statistician*, 1988. **42**(1): p. 59-66.
- [21] Wagemans, J., et al., *A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization*. 2012.